SC-RR-69-310

A MODEL FOR THE QUANTIFICATION OF THE QUALITATIVE MICROBIAL SAMPLING PROBLEM

A. L. Roark

Planetary Quarantine Department
Sandia Laboratories
Albuquerque, New Mexico 87115

May 1969

This work was conducted under Contract No. W-12,853, Bioscience Division, Office of Space Science Applications, NASA Headquarters, Washington, D. C.

TABLE OF CONTENTS

Introduction

In the design of an information system for lunar programs for the Planetary Quarantine Officer of NASA [7], there is described a mathematical model called the "Probability-of-Identifying-All Model". This model is one which estimates the probability that all species of microorganisms on an Apollo module have been identified by the environmental and surface sampling done by the Spacecraft Bioassay Laboratory of the Public Health Service at Cape Kennedy. This document describes a simple model which could be used for this purpose. It will also provide information which will be useful to the Planetary Quarantine Officer in establishing sampling protocol.

Many works in statistics have approached the problem of estimating the unknown number N of mutually exclusive classes (species of microorganisms) into which a population of size M (number of microorganisms) is divided, but none are entirely applicable to the model we have just described. Goodman [3] has considered this problem when the total population size M is known and is finite. He also assumes that the maximum number of elements in each of the N classes is known. For his work to be applicable to Apollo is impossible because of the biologically uncontrolled environments which the Apollo modules see at Cape Kennedy and because the limit obtained by letting these numbers become unbounded is not within the scope of his work. This work could prove useful at a later date in laminar flow situations where the total number of microbes is controlled and may be known.

Another form of a solution to this problem is sequential sampling methods. Several authors have looked at these methods [1,4]. Sequential

methods have the disadvantage that they are not useful in the determination of sampling protocol since the number of samples needed will change as sampling proceeds. One of these methods may be useful in the applications described in the information system design.

Obviously, we could not hope for a model which would determine, in advance of the actual sampling, the number of samples needed, independent of the number of species of microorganisms we expect to find. Thus, this document will outline a model which, even though it is not all that is desired, is simple to understand and which does provide the information needed in establishing sampling protocol.

The Model

Let N represent the number of species of microorganisms which are on a given surface or in a given environment. Let us suppose we sample one microorganism at a time and identify the colony which it produces. Suppose we record this observation. What we shall first attempt to calculate is the probability $P_{n,N}$ that all N species of microorganisms have been observed after identifying n colonies. In order to do this we assume that the number of microorganisms of each specie is large enough that this sampling without replacement is approximated by sampling with replacement. Since we do not have any better information, we will also assume that equal numbers of each specie of microorganism exist in the environment. The author is well aware of the fact that these assumptions may be open to challenge but feels that they are approximations to the actual situation which yields useable results.

Let us consider first the event that one specie has not been identified after identifying n colonies. The n colonies are chosen at random and must be divided among the N-1 remaining species. Since all N species are

are equally likely, this can be done in $(N-1)^n$ different ways. If two species go unidentified, then the observed colonies divide themselves among the N-2 other species in $(N-2)^n$ ways, etc. Since the total number of ways in which the n colonies are divided among the N species is $(N)^n$ it is easy to see that the probability that any one specie goes unidentified is

$$P_1 = \binom{N}{1}(1-1/N)^n.$$

Generalizing this we see that the probability that any k species have gone unidentified after sampling n colonies is

$$P_k = \binom{N}{k}(1-k/N)^n, \quad k = 1,2, \ldots .$$

Thus

$$P_{n,N} = \sum_{k=0}^{N} (-1)^k \binom{N}{k}(1-k/N)^n. \qquad (1)$$

If N is known, this provides the desired probability. Through their own experimentation and that of others, the Public Health Service already has a good estimate of the number of species it expects to find, and since there is a finite (even though constantly changing) number of species, which have been identified by microbiologists around the world, we shall assume that N is known. Observe that as N increases $P_{n,N}$ decreases and thus a conservative estimate of N is best.

Before going on to an application of this model in the determination of the number of colonies which we must observe in order to estimate the

qualitative bioburden on an Apollo module, it is necessary to point out that equation (1) is not new, but arises in "ball in urn problems" [1,2].

Sampling Procedures

Suppose we wish to determine the smallest number of colonies $\tilde{n}$ for which $P_{\tilde{n},N} \geq \alpha$, where α is preassigned and $0 < \alpha < 1$. Feller [2] shows that if $N,n \to \infty$ in such a way that $Ne^{-n/N} \to \lambda$ where $0 < \lambda < \infty$, then

$$P_{N,n} \to e^{-\lambda}. \tag{2}$$

If N is large, this yields the result that

$$\tilde{n} = N \ln N + \gamma N + O(N) \tag{3}$$

where γ is determined by the equation $e^{-e^{-\gamma}} = \alpha$.

Our approach to the determination of sampling procedures has been to select representative values for N and α and using the GE 435 to compute corresponding values of $\tilde{n}$. (Other researchers may elect to keep another combination of two of the three parameters fixed and solve for the third).

Table 1 gives some of the values of γ for some typical values of α computed using Newton's method. This will be useful for those wishing to perform other calculations. Table 2 gives $\tilde{n}$ for various choices of N and α.

Table 1 - Solutions of $e^{-e^{-\gamma}} = \alpha$

α	γ
.50	.366513
.70	1.03093
.80	1.49994
.90	2.25037
.95	2.9702
.99	4.60015

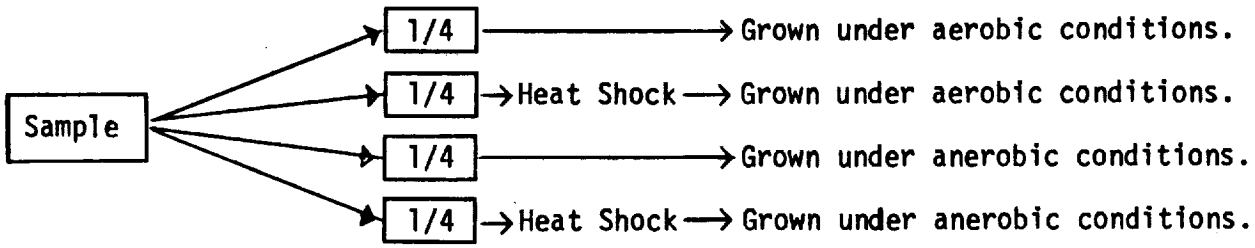
Table 2 - Values of $\tilde{n}$ for Fixed Values of N and α

N\α	.50	.70	.80	.90	.95	.99
50	214	248	271	309	345	426
100	498	564	611	686	758	921
200	1133	1266	1360	1510	1654	1980
500	3291	3623	3858	4233	4593	5408
1000	7275	7939	8408	9159	9878	11508
1500	11520	12517	13220	14346	15426	17871
2000	15935	17264	18202	19703	21143	24403

Conclusions

Throughout this paper, the sampling space which we have used is the set of microorganisms in the immediate vicinity of the Apollo modules which can be identified by the procedures used by the Public Health Service at the Spacecraft Bioassay Laboratory at Cape Kennedy [5]. There are microorganisms which are known to exist and which cannot be identified using their methods. Whether these microorganisms are present in the areas surrounding the Apollo modules is unknown since no one has attempted to identify them in that environment. This class of microorganisms is probably small and most of the microorganisms in which the Planetary Quarantine Officer is interested in identifying are not excluded from our sample space because of the procedures used to identify the colonies.

Aspects of the protocol used by the Spacecraft Bioassay Laboratory which will effect the conclusions reached by our model are basic in their handling of the samples. We can depict the important parts of their procedures for this discussion schematically as follows:



The belief has been expressed that the qualitative (and perhaps the quantitative) properties of a sample can be changed during ultrasonication and heat shocking. It could also happen that the incubation conditions could change the qualitative properties of microorganisms. Since quantitative information about the qualitative changes which take place as a result of these steps is not available we shall omit them for now.

The purpose of the heat shock and the different conditions under which the sample is incubated are to prevent colonies from growing unless they fall into a particular category. The assumption which we have made covering these steps implies that this is what is accomplished. If we also assume that the division of the sample into four parts is exact, then the probability of a colony growing once it has been removed from the spacecraft is 1/4. Assuming, as we have, that all species are equally likely and that the sampling is random, the probability of a microorganism being removed and growing into a colony is equally likely for every specie. Thus, our sample space (which are those that can be removed and identified) is the same as those which can be removed, and the probability of this is equal for every specie in which we are interested. This satisfies the conditions of our model.

From private conversations with representatives of the Public Health Service at Cape Kennedy [6], we find that at the present sampling rates on the various modules, they identify between four and five thousand colonies on each module. This means that they are capable of identifying 400 to 500 species with a probability of .95. If they feel that there are only 300 species of microorganisms which they wish to identify and if they wish to identify these species with probability .80 as has been mentioned by some, this means they can reduce their qualitative identification by fifty percent.

If one is conservative and assumes that all species of microorganisms capable of growth with current protocol exist in the areas surrounding the Apollo modules, then N may be between 1500 to 2000. This would mean an increase of four to five times the present identification rate in order to work at 99% confidence. Under the present sampling protocol they are operating at less than 50% confidence for this number of species which is probably not satisfactory.

References

1. Darling, D. A.and H. Robbins, "Finding the Size of a Finite Population", unpublished manuscript.
2. Feller, W., Introduction to Probability Theory and Its Applications, Vol. 1, 2 ed., John Wiley & Son, Inc., New York (1957).
3. Goodman, L. A., "On the Estimation of the Number of Classes in a Population", Ann. Math. Stat., Vol. 20 (1949), pp. 572-579.
4. Goodman, L. A., "Sequential Sampling Tagging for Population Size Problems", Ann. Math. Stat., Vol. 24 (1953), pp. 56-69.
5. National Aeronautics and Space Administration, Standard Procedures for the Microbiological Examination of Space Hardware, NASA Publication, 1966.
6. Puleo, J. R., private communication, August 6, 1968.
7. Roark, A. L. and A. L. Wyer, "An Interactive Computer Information System for Planetary Quarantine for Lunar Programs", Sandia Laboratories Publication SC-RR-68-545, July, 1968.

DISTRIBUTION:
J. A. Hornbeck, 1
D. R. Cotter, 1000
R. T. Dillon, 101
B. H. VanDomelen, 101
W. J. Howard, 1000
W. A. Gardner, 1500
H. E. Lenander, 1600
T. M. Burford, 1700
C. Winter, 1710
D. P. Peterson, 1711
M. J. Norris, 1720
J. M. Worrell, Jr., 1721
R. G. Clem, 1730
H. D. Sivinski, 1740
A. L. Roark, 1741 (6)
R. W. Henderson, 2000
L. D. Smith, 2300
C. B. McCampbell, 2310
J. W. Jones, 2350
L. J. Paddison, 2400
R. S. Claassen, 2600
J. L. Rodgers, 2630
J. R. Sublett, 2640
R. B. Powell, 3000
C. W. Campbell, 4000
S. J. Buchsbaum, 5000
L. C. Hebel, 5200
C. F. Bild, 5400
D. W. Ballard, 5425
R. A. Bice, 7000
L. E. Lamkin, 7300
L. J. Heilman, 7400
C. R. Barncord, 7600
G. A. Fowler, 9000
J. H. Scott, 9200
A. Y. Pope, 9300
L. A. Hopkins, Jr., 9500
J. W. McKiernan, 9521
W. F. Carstens, 3410
L. C. Baldwin, 3412
W. J. Wagoner, 3413
B. R. Allen, 3421
B. F. Hefley, 8232

NASA, Code SC
Grants and Contracts
400 Maryland Avenue, S.W.
Washington, D. C. 20546 (25)

L. B. Hall, NASA
Code SB
400 Maryland Avenue, S.W.
Washington, D. C. 20546 (25)

E. D. Smyth
Acting Chief
Special Projects Branch
Sandia Area Office
U. S. Atomic Energy Commission
Albuquerque, New Mexico 87115

S. A. Upson
Director, Non-Weapons Activity Div.
Office of Plans and Budgets
U. S. Atomic Energy Commission
Albuquerque, New Mexico 87115

University of California, LRL
P. O. Box 808
Livermore, California 94551
Attn: Tech. Info. Div.
For: Report Librarian

Los Alamos Scientific Laboratory
P. O. Box 1663
Los Alamos, New Mexico
Attn: Report Librarian

Richard G. Bond
School of Public Health
College of Medical Science
University of Minnesota
Minneapolis, Minnesota 55455

John H. Brewer
Star Route 2
Brownwood, Texas 76801

Harold Walker
Director of Research Services
Graduate College
University of New Mexico
Albuquerque, New Mexico

Frank B. Engley, Jr.
Chairman, Department of Microbiology
School of Medicine
University of Missouri
Columbia, Missouri

Gilbert V. Levin
Hazleton Laboratories, Inc.
Box 30
Falls Church, Virginia

Irving J. Pflug
Professor of Environmental Health
University of Minnesota
College of Medical Sciences
Minneapolis, Minnesota 59455

Gerald Silverman
Department of Nutrition and Food Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

John A. Ulrich
Department of Microbiology
Mayo Clinic
Rochester, Minnesota 55902

Samuel Schalkowsky
Exotech Incorporated
525 School Street, S.W.
Washington, D. C. 20024

Charles Draven
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, California 91103

Mark A. Chatigny
Research Engineer
Naval Biological Laboratory
Naval Supply Center
University of California, Berkeley
Oakland, California 94625

Richard G. Cornell
Associate Professor of Statistics
Department of Statistics
Florida State University
Tallahassee, Florida

Martin S. Favero
Department of Health, Education
and Welfare
CDC-Phoenix Field Station
4402 North 7th Street
Phoenix, Arizona 85014

F. N. LeDoux
Head, Structural and Mechanical
Applications Section
Goddard Space Flight Center
Greenbelt, Maryland

Q. Ussery
Code NC3, Quality Assurance Branch
Manned Spacecraft Center, NASA
Houston, Texas

F. J. Beyerle
George C. Marshall Space Flight Center
Manufacturing Engineering Laboratory
Code R-ME-MMC
Huntsville, Alabama 35812

J. Gayle
Code SOP
Kennedy Space Center, NASA
Cape Kennedy, Florida

E. Rich
Code 624
GSFC Sterilization Laboratory
Goddard Space Flight Center
Greenbelt, Maryland 20771

N. H. MacLeod
Space Biology Branch
Code 624, Bldg. 21, Rm. 161
Goddard Space Flight Center
Greenbelt, Maryland 20771

Robert Angelotti
Deputy Chief, Milk and Food Research
Robert A. Taft Sanitary Engineering
Center
Cincinnati, Ohio

Erwin Fried
Rm. M-2101
General Electric Company
P. O. Box 8555
Philadelphia, Pennsylvania 19101

Martin G. Koesterer, Microbiologist
Bioscience Operation
General Electric
P. O. Box 8555
Philadelphia, Pennsylvania 19101

Carl Bruch
Chief, Bacteriology Branch
Division of Microbiology
Food and Drug Administration
3rd & C., S.W.,Room 3876
Washington, D. C. 20204

John W. Beakley
Department of Biology
University of New Mexico
Albuquerque, New Mexico

Loren D. Potter, Chairman
Department of Biology
University of New Mexico
Albuquerque, New Mexico

Loris W. Hughes
Department of Biology
New Mexico State University
University Park, New Mexico

Joe Stern
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, California 91103

J. J. McDade
Biohazards Group
Pitman-Moore Company
Dow Chemical Company
P. O. Box 1656
Indianapolis, Indiana 46206

William K. McKim
Aerospace Corporation
Building A2, Room 2019
2350 East El Segundo Blvd.
El Segundo, California

Edwin J. Sherry
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, California 91103

James A. Green
Life Sciences Systems (BB14)
North American Rockwell Corp.
12214 Lakewood Blvd.
Downey, California

Fred L. Whipple
Smithsonian Astrophysical Observatory
Cambridge, Massachusetts 02138

Arthur H. Neill
Code SB
400 Maryland Avenue, S.W.
Washington, D. C. 20546

Rudy Puleo
Public Health Service
Spacecraft Bioassay Laboratory
Drawer Y
Cape Kennedy, Florida 32900

USAEC
Division of Technical Information
P. O. Box 62
Oak Ridge, Tennessee 37830
Attn: Reference Branch
P. E. Postell

Carl Sagan
Cornell University
Center for Radiophysics and Space
Research
Space Science Building
Ithaca, New York 14850

Document Library
Lovelace Foundation for Medical
Education and Research
5200 Gibson Blvd. SE
Albuquerque, New Mexico 87108

Martin S. Tierney
Group J-10
Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87544

Virgil L. Dugan
University Village
Apartment 28-K
Starkville, Mississippi 39759

Lawrence P. Chambers
NASA Headquarters
Office of Manned Space Flight
Code MLR
Washington, D. C. 20546

Cloyce E. Matheny
Section Manager
Planetary Quarantine
Space and Missile Division
McDonnell Aircraft
St. Louis, Missouri

L. P. Daspit, Jr.
Viking Project Quarantine Officer
Viking Project Office
Langley Research Center, NASA
Hampton, Virginia 23365